Original Article

Variant analysis of the first Lebanese SARS-CoV-2 isolates

Mhamad Abou-Hamdan [a], Kassem Hamze [a], Ali Abdel Sater [b], Haidar Akl [a], Nabil El-zein [a], Israa Dandache [a], Fadi Abdel-sater [c,*]

[a] Biology department, Faculty of sciences (I), Lebanese University, Beirut, LB, Lebanon
[b] Faculty of Medicine, Lebanese University, Beirut, LB, Lebanon
[c] Biochemistry department, Faculty of sciences (I), Lebanese University, Beirut, LB, Lebanon

ABSTRACT

Recently the first genome sequences for 11 SARS-CoV-2 isolates from Lebanon became available. Here, we report the detection of variants within the genome of these strains. Pairwise alignment analysis using blastx was performed between these sequences and the UniProtKB data for the SARS-CoV-2 coronavirus to identify amino acid variations. Variants analysis was performed using multiple Bioinformatics tools. We noticed for the first time 18 mutations that have never been reported before. Among those, a frame shift (8651A>) in NSP4, a stop codon 6887A > T in NSP3 and two missense mutations in spike S2 were found. In addition, we found 28 variants in ORF1ab alone. A previously reported variant, 23403A > G, in the spike protein S2 was mostly seen. Two other known mutations 25563G > T in ORF3a and 14408C > T in ORF1ab were detected respectively in 6 and 8 out of the 11 isolates. Our results may help to prognose forthcoming infections in this region.

1. Introduction

The Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-Cov-2) is a newly identified β-coronavirus that was declared as a pandemic by the World Health Organization (WHO) on March 11, 2020. SAR-CoV2 is an enveloped, positive sense RNA virus that was found in humans and other mammals [1,2]. Phylogenetic analysis of 160 complete genomes of SARS-Cov-2 isolated from different countries, identified three central variants (A, B, and C) based on variabilities in their amino acid sequence [3]. Although such analysis helped in tracing routes of infections, we still need an important number of cases to be studied and compared [3].

The genome of SARS-Cov-2 is composed of a total of 11 genes with 11 open reading frames (ORFs)0. ORF1ab after proteolytic cleavage encodes 16 nonstructural proteins (NSPs); S, E, M and N, encode structural proteins, namely spike, envelope, membrane and nucleocapsid proteins; and ORF3a, ORF6, ORF7, ORF8 and ORF10 encode accessory proteins [1,2]. The latter protein group is known to have high variability in its sequence [3].

ORF1ab consists of two third of the entire genome and produces two polyproteins, pp1a and pp1ab, these are cleaved into 16 nonstructural proteins (NSP 1 - NSP 16) [1]. NSP 3 is characterized to be a multifunctional protein, acting as viral protease and able to suppress interferon responses [1]. The NSP 12 is the RNA-dependent RNA polymerases (RdRps) of SARS-CoV-2, this is the major enzyme in the replication and transcription of the viral genome [4]. NSP 14 also named ExoN, is reported to have an exonuclease activity with proofreading function; it has been reported that any variant in this protein will make the genome of this SARS-CoV-2 strain prone to high mutational changes [4]. Moreover, scientists proved the association of NSP 7 and NSP 8 with NSP 12, forming a stable supercomplex [4]. Altogether these proteins will ensure the transcription fidelity of the virus. ORF2 encodes the spike proteins (S) [2]. This glycoprotein is divided into two subunits, a globular S1 subunit that contains the receptor-binding domain (RBD) whereas S2 subunit has the domain involved in fusion [1]. It was well demonstrated that through receptors on the host cell-membrane the spike proteins mediate the entry of SARS-Cov-2 into human cells [3]. Mutations affecting the S protein will induce perturbations in the virus entry. Therefore, these glycoproteins were the target of neutralizing antibodies investigations. Even though, structural proteins, being mostly exposed to the host immune response, grabbed all the attention of therapeutic research, yet nonstructural components play as well a major role in virus-host interactions [2]. Out of this group, ORF3a encodes an ion channel protein related to inflammasome, inducing the activation of caspase 1 and the maturation of IL-1β [2]. It was lately described that SARS-CoV-2 induces cytokine storm and pyroptosis of the host cells

---

which in many cases end in severe symptoms leading to death [5].

The 2020 pandemic caused by SARS-CoV-2 still lacks approved countermeasures to control the infection. Despite the efforts for data collection worldwide to control the spread, scientists are still trying to understand the genome variability and evolution of this virus. The purpose of this study is to investigate the different variants found in the Lebanese isolates infected with SARS-CoV-2. Our data may provide new evidence to understand the etiology of the variation in symptoms exacerbation among populations.

2. Material and methods

The Lebanese suspected cases between February and March 2020 were collected by the national reference hospital at Lebanon. Diagnostic purposes for the presence of SARS-CoV-2 were followed as per the guidelines of the Lebanese Ministry of Health.

For this study, we obtained 11 complete genomic sequences of Lebanese SARS-CoV-2 isolates from the GISAID's EpiCoVTM Database (https://www.gisaid.org/). Among them, eight sequences were obtained from patients with a travel history to SARS-CoV-2-affected countries, namely Egypt (1 sample), Iran (3 samples), United Kingdom (2 samples), France (1 sample), Italy (1 sample) and the remaining three sequences from local residents with no travel history, at least, for the past three months ahead of their infection. Even though some of the isolates had low-quality sequence reads described by a stretch of a single consecutive nucleotide, but variants detection was possible. The accession numbers of the isolate strains used in this study are EPI_ISL_454420, EPI_ISL_450508 to EPI_ISL_4505017.

We used the Wuhan (hu-1) virus sequence as a genome reference (NCBI Genbank, NC_045512) and the following sequences as a protein SARS-CoV-2 reference (New prortal Covid19 uniprot: PODTD1, PODTC3, PODTC4, PODTC4, PODTC5, PODTC6, PODTC7, PODTC8, PODTC9, PODTD2, PODTD3, PODTD8).

Mutations specific to the Lebanese SARS-CoV-2 isolates were identified using multiple analysis. First of all, we identified amino acid mutations in Lebanese isolates by extracting pairwise alignment to each reference protein (downloaded from Covid19 Uniprot) using Blastx. Then, we analyzed sequence variations using Clustal Omega tool that conduct multiple sequence alignment between the genome of Lebanese isolates and the reference genome (NC_045512). the sequence variations analysis was also performed using VipR analysis tool and coVsurver enabled by GISAID. This data was checked and validated carefully

against aligned sequences. Finally, we investigated the mutation existence and the global frequency from the previously reported worldwide data of SARS-CoV-2, by using CoVsurver enabled by GISAID (based on viral sequences in GISAID's EpiCoV database) and SARS-CoV-2 Mutation Browser v-1.3, (http://covid-19.dnageography.com/).

Two tables were generated showing the identified variants. Variations in sequences showing ambiguity were noted as not determined (ND).

3. Results

A 67 total variants were found with 40 unique missense variants as shown in Tables 1 and 2. For what it concerns the 40 missense variants, 28 were found in ORF1ab, which is the longest ORF occupying 2/3 of the entire genome. ORF1ab is cleaved into many nonstructural proteins (NSP1-NSP16). Among them, NSPs, NSP3 and NSP4 had the highest number of variants in the analyzed samples. The most common variants were 23403A > G in the spike protein S2 and 14408C > T in ORF1ab, both in eight samples, and 25563G > T in ORF3a in six samples (Table 1). The occurrence of these three mutations coincided in the three local patients and three travelling patients as indicated in Table 2.

Interestingly, we noticed for the first time 18 mutations that have never been reported before in the literature or found in Gisaid (containing sequence analysis of 95,522 SARS-CoV-2 strains) or in the SARS-CoV-2 Mutation Browser v-1.3 (containing sequence analysis of 10,416 SARS-CoV-2 strains from 105 locations) (Table 1). Among these mutations, four novel mutations were found only in the local patients (NSP3: 6198C > A; NSP8: 12297A > T; NSP4: 8651A > fs; ORF3a: 26103-5TGA > del), two were found in local patients and in patient travelling from France (NSP3: 6281A > G/T; 6285C > A). 12 novel mutations were identified in travelling patients: one stop codon (6887A → T) in NSP3 (Table 2), eight mutations in ORF1ab: NSP3(7766A > C), NSP4 (8897A > T), NSP5(10,595 T > C), NSP12 (14369G > T, 14993C > T) NSP13 (16301G > T) and NSP14(18670G > T, 19499A > C), two missense mutations in spike S2 (22093G > T, 22425C > T) and one mutation in E gene (26428A > T).

The patient travelling from Egypt died correlating the deterioration of his clinical status with a past medical history and old age. While, the sequence of this strain is 100% identical with the strain isolated from a local old released patient (EPI_ISL_450511). Finally, one of the two patients travelling from Iran, isolated on the 27th of February, showed identical genome sequence as the reference strain with no identified

Table 1
Variants detected in coding sequences of 11 Lebanese SARS-CoV-2 isolates.

Accession	Date	Travelling from	ORF1ab	Spike	ORF3a	E	ORF7a	Nucleoprotein
EPI_ISL_450508	04–03-20	Egypt	P4715L	D614G	Q57H	–	–	–
EPI_ISL_450510	04–03-20	Iran	Q22R, R207C, V378I, L3606F	ND	–	ND	–	–
EPI_ISL_450512	27–02-20	Iran	R207C, V378I, A540V, I2501V, I2511V, M2796I, L3606F	–	–	–	–	–
EPI_ISL_450520	21–02-20	Iran	S2797F, K2798R, C4703F, P4715L, R5346I, D6136Y, H6412P	D614G	Q57H	–	–	–
EPI_ISL_450511	04–03-20	UK	P4715L	D614G	–	–	–	R203K, G204R
EPI_ISL_450515	11–03-20	UK	R207C, V378I, M2796I, L3606F	M153I	–	–	–	–
EPI_ISL_450513	09–03-20	France	N2006V, T2007N, K2208stop, E2089D, M2796L, F3444L, P4715L	L54F, A287V, D614G	–	V62D	–	R203K, G204R
EPI_ISL_450517	09–03-20	Italy	A1938V, M2796I, S2797F, K2798R, N2878Y, A3270V, P4715L, S4910L	M177R, D614G	Q57H	–	P84S	–
EPI_ISL_450514	13–03-20	Local	S1978Y, Frame shift 8651A>, P4715L	D614G	Q57H, L73F	–	–	–
EPI_ISL_450516	15–03-20	Local	N2006V, T2007N, L3199R, Q4012L, P4715L	D614G	Q57H, D238	–	–	–
EPI_ISL_450509	15–03-20	Local	P4715L	D614G	Q57H	–	–	–

Table 2
Variants frequencies of Lebanese SARS-CoV-2 isolates.

Gene Names	Nucleotide variation	Amino acid change	Lebanese sample #	Frequency in GISAID	GISAID sample #
ORF1ab/ NSP1	330A > G	Q22R	1	0%	2
ORF1ab/ NSP2	884C > T	R207C	2	0.2%	192
	1397G > A	V378I	2	0.54%	518
	1884C > T	A540V	1	0.09%	85
ORF1ab/ NSP3	6078C > T	A1938V	1	0%	2
	6198C > A	S1978Y	1	0%	1*
	6281A > G/T	N2006V	2	0%	2*
	6285C > A	T2007N	2	0.0d%	2*
	6532G > T	E2089D	1	0.04%	41
	6887A > T	K2208stop	1	0%	1*
	7766A > C	I2501V	1	0%	1*
ORF1ab/ NSP4	8651A>	Frameshift	1	0%	1*
	8651A > C	M2796L	3	0.08%	78
	8653G > T	M2796I	1	0.18%	175
	8655C > T	S2797F	2	0.08%	75
	8658A > G	K2798R	2	0.02%	16
	8897A > T	N2878Y	1	0%	1*
	9860 T > C	L3199R	1	0.01%	3
ORF1ab/ NSP5	10074C > T	A3270V	1	0.01%	8
	10,595 T > C	F3444L	1	0%	1*
ORF1ab/ NSP6	11,083 T > G	L3606F	1	9.59%	9150
ORF1ab/ NSP8	12297A > T	Q4011L	1	0%	1*
ORF1ab/ NSP12	14369G > T	C4703F	1	0%	1*
	14408C > T	P4715L	8	79.44%	75,899
	14993C > T	S4910L	1	0%	1*
ORF1ab/ NSP13	16301G > T	R5346I	1	0%	1*
ORF1ab/ NSP14	18670G > T	D6136Y	1	0%	1*
	19499A > C	H6412P	1	0%	1*
Spike	21724G > T	L54F	1	0.01%	2
	22021G > T	M153I	1	0.07%	67
	22093G > T	M177R	1	0%	1*
	22425C > T	A287V	1	0%	1*
	23403A > G	D614G	8	79.71%	76,294
ORF3a	25563G > T	Q57H	6	22.91%	21,884
	25611C > T	L73F	2	0.01%	14
	26,103-5TGA>	D238del	1	0%	1*
E	26428A > T	V62D	1	0%	1*
ORF7a	27643C > T	P84S	1	0.02%	20
Nucleoprotein	28881G > A	R203K	2	34.88%	33,301
	28883G > C	G204R	2	34.77%	33,199

* : *Reported for the first time; #: Number; GISAID sample #: 95522 reported strains on 15 september 2020.*

mutations in spike protein, but 7 mutations in ORF1ab. We expect out of our results that these haphazard described mutations will be detected from new Lebanese SARS-CoV-2 isolates and will most likely spread after gradual reopening normal life.

4. Discussion

Comparative analysis of genome sequences of SARS-CoV-2, from different worldwide isolates, revealed new variants that could be involved in varied exacerbation of symptoms in patients. Here we describe 40 unique missense variants investigated from 11 sequenced SARS-CoV-2 isolates, identified from individuals being present in Lebanon when presented to the hospital with SARS-CoV-2 symptoms. Even though the number of the sequenced isolates is very low and just 3 out of the 11 were local residents, it was very important to study the incidence and frequency of such important number of mutation(s) rarely seen in other countries.

We observed the emergence of 18 novel mutations from only 11 sequences, not described before in the literature or in any database. Of those 18 mutations, sixteen missense mutations were found in different positions as follows; 6198C > A, 6281A > G/T, 6285C > A, 6887A > T and 7766A > C belonging to NSP 3, 8897A > T belonging to NSP 4, 10,595 T > C belonging to NSP 5, 12297A > T belonging to NSP 8, 14369G > T and 14993C > T belonging to NSP 12 (RdRp), 16301G > T belonging to NSP 13, 18670G > T and 19499A > C belonging to NSP 14, 22093G > T and 22425C > T belonging to S and 26428A > T belonging to E gene. This was expected since after rapid spread of the virus across countries, many reports were published describing mutation hotspots, and correlating the results to the variable clinical condition of Covid-19 patients among populations [4]. Most of our newly described mutations belong to non-structural proteins (NSPs). Among these, we identified five distinctive mutations in ORF1ab affecting the NSP 3 protein (Table 1). This protein was described previously to suppress host interferon response and to interact with other protein playing a role in viral replication. In one study, assessing mutations in non-structural viral proteins, the authors considered that NSP 3 protein might loses its stability upon specific mutational changes [6].

Moreover, we reveal the identification of two extremely important novel mutations, a stop codon and a frameshift mutation affecting NSP 3 and NSP 4 at position 6887A > T and 8651/fs, respectively. The onset of such deleterious mutation could be explained by the onset of the mutation after the entry of the virus to the human body [7]. It is well known that RNA viruses mutate at a very high rate, with a possible appearance of mutations in a patient every day, as it was reported for HIV virus infection [4]. Usually viruses use such modalities to adapt and survive antiviral therapies, but sometimes it fails to regulate such mechanism producing deleterious genome modifications [7]. Nevertheless, such types of mutation, frame shift and codon stop, were previously reported in essential genes, and alternative proteins expression strategies were proposed, such as ribosomal frameshifting and shunting [8,9]. It seems obvious that subsequent laboratory work is needed to confirm the biological activity of the mutated strain. In addition, the analyzed sequence must be representative of the viral population infecting the patient, and not selective, in case of a double population with a non-mutated strain complementing the mutated one. Double viral population was documented in a recent study, where the authors reported the presence of 29 nucleotides deletion segment in the gene coding for accessory protein ORF 9 that eliminated additionally ORFs 10 and 11. They claimed to detect a co-existence of a non-deleted genome and the deleted one in the same sample from the same patient [10]. In our results, we detected a frequent missense mutation among 6 of our 11 isolates (25563G > T), and another deletion mutation belonging to the gene of the accessory protein ORF3. The ORF3a induces apoptosis and inflammatory responses in the infected cells [11]. ORF3a protein contains TNF receptor-associated factors (TRAF), ion channel, and caveolin binding domain. Near to these three domains is located the Q57H mutation, which may affect the inflammasome activation [11].

Many previously characterized mutations were also observed in our work. Out of these, we identified mutations in the structural Spike coding gene [8,12]. We found a major mutation reported 76,294 times in GISAID database at position 23403A > G (D614G). Korber et al. have

shown that this variant is associated with greater infectivity as well as clinical evidence that it is associated with higher viral loads but does not appear to cause a more serious form of disease [13]. Two less frequently reported at positions 21724G > T and 22021G > T; and two additional novel mutations based on our analysis at positions 22093G > T, 22425C > T [12]. It is well known, that any mutation at the level of S gene, coding for one of the structural proteins, will affect the attachment and the transmission of the disease [3]. Among structural proteins, we describe the presence as well of two mutations affecting the N gene in two positions, 28881G > A and 28883G > C, in two isolates from patients coming from France and UK.

Of high interest in our variants was the co-existence of 3 mutations (23403A > G, 14408C > T and 25563G > T) in 8 out of the 11 sequences (with only two strains, lacking the 25563G > T mutation). This combination of changes affected simultaneously, at one place, the nonstructural, structural, and accessory proteins. Such occurrence was statistically proved and correlated to the presence of the RdRp mutation at position 14408C > T [4]. When found, this mutation presumably affecting the proofreading activity of RdRp, provokes the onset of other changes [4]. At the clinical level, we noticed that two patients having the same age, one travelling from Egypt and the other is among local residents, had identical changes harboring the previous association of mutations but with different prognosis. The patient coming from Egypt died, correlating the deterioration of his health with his past medical history. Additional random mutation combinations mainly at the level of NSPs were found in three patients travelling from Iran and UK, namely mutations at position 884C > T; 1397G > A; 8653G > T and 11,083 T > G.

This work gives a first observation and comparative analysis of variants isolated from this region. More genome sequences of isolates from local residents are needed to establish a complete evolutive study of the virus in our region. Our results pave the way for predicting the type of strains that would most likely spread in Lebanon, mainly after reopening normal life. This would most likely lead to a better and accurate vaccine generation before another wave of coronavirus spread.

5. Conclusion

In conclusion, more investigation is needed to study the clinical significance of these different types of mutations and other possible ones. Our analysis of SARS-CoV-2 genomes from Lebanese isolates, obtained at the beginning of the virus spread in this region, has revealed novel type of mutations, up to our knowledge not yet described in any database. Nonetheless, we here consider that the frameshift and the stop codon mutations in ORF1ab would not change in the expression of SARS-CoV-2, particularly taking into consideration the alternative unconventional expression mechanisms of viruses. Such type of deleterious mutations in the viral genome should be monitored as they could significantly modify symptoms severity.

Declaration of Competing Interest

References

[1] D. Forni, et al., Molecular evolution of human coronavirus genomes, Trends Microbiol. 25 (1) (2017) 35–48.

[2] F.K. Yoshimoto, The Proteins of Severe Acute Respiratory Syndrome Coronavirus-2 (SARS CoV-2 or n-COV19), the Cause of COVID-19, Protein J **39** (3) (2020) 198–216.

[3] R.J. Hulswit, C.A. de Haan, B.J. Bosch, Coronavirus spike protein and tropism changes, Adv. Virus Res. 96 (2016) 29–57.

[4] M. Pachetti, et al., Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant, J. Transl. Med. 18 (1) (2020) 179.

[5] M.Z. Ratajczak, M. Kucia, SARS-CoV-2 infection and overactivation of Nlrp3 inflammasome as a trigger of cytokine "storm" and risk factor for damage of hematopoietic stem cells, Leukemia 34 (7) (2020 Jul) 1726–1729.

[6] D. Benvenuto, et al., Evolutionary analysis of SARS-CoV-2: how mutation of non-structural protein 6 (NSP6) could affect viral autophagy, J. Inf. Secur. 81 (1) (2020) e24–e27.

[7] J.M. Cuevas, et al., Extremely High Mutation Rate of HIV-1 In Vivo, PLoS Biol **13** (9) (2015) e1002251.

[8] T. Koyama, D. Platt, L. Parida, Variant Analysis of COVID-19 Genomes, 2020.

[9] F.D. Ramos, et al., Programmed −1 ribosomal frameshifting in the SARS coronavirus, Biochem. Soc. Trans. 32 (6) (2004) 1081–1083.

[10] J.W. Tang, et al., The large 386-nt deletion in SARS-associated coronavirus: evidence for quasispecies? J. Infect. Dis. 194 (6) (2006) 808–813.

[11] S.S. Hassan, et al., Molecular conservation and differential mutation on ORF3a gene in Indian SARS-CoV2 genomes, Genomics 112 (5) (2020) 3226–3237.

[12] J.S. Kim, et al., Genome-wide identification and characterization of point mutations in the SARS-CoV-2 genome, Osong Public Health Res Perspect 11 (3) (2020) 101–111.

[13] B. Korber, et al., Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus, Cell 182 (4) (2020), 812–827 e19.